

ToxoDB: an integrated *Toxoplasma gondii* database resource

Bindu Gajria¹, Amit Bahl¹, John Brestelli², Jennifer Dommer², Steve Fischer²,
 Xin Gao², Mark Heiges³, John Iodice², Jessica C. Kissinger^{3,4,*},
 Aaron J. Mackey¹, Deborah F. Pinney², David S. Roos¹, Christian J. Stoeckert Jr²,
 Haiming Wang³ and Brian P. Brunk²

¹Department of Biology, ²Department of Genetics, Center for Bioinformatics, University of Pennsylvania, Philadelphia, PA, ³Center for Tropical & Emerging Global Diseases and ⁴Department of Genetics, University of Georgia, Athens, GA, USA

Received September 17, 2007; Revised October 17, 2007; Accepted October 18, 2007

ABSTRACT

ToxoDB (<http://ToxoDB.org>) is a genome and functional genomic database for the protozoan parasite *Toxoplasma gondii*. It incorporates the sequence and annotation of the *T. gondii* ME49 strain, as well as genome sequences for the GT1, VEG and RH (Chr Ia, Chr Ib) strains. Sequence information is integrated with various other genomic-scale data, including community annotation, ESTs, gene expression and proteomics data. ToxoDB has matured significantly since its initial release. Here we outline the numerous updates with respect to the data and increased functionality available on the website.

INTRODUCTION

Toxoplasma gondii is an intracellular apicomplexan parasite capable of infecting humans. Infection is typically asymptomatic in healthy individuals, but may lead to congenital birth defects and encephalitis in immunosuppressed individuals (1,2). ToxoDB, initially released in May 2001, has been substantially updated in both content and functionality since last described in January 2003 (3). ToxoDB provides access to the genome sequence and annotation of the *T. gondii* ME49 strain. It also incorporates the genomic sequence of multiple other strains. The parasite genome is ~63 Mb in size and consists of 14 chromosomes (4).

The initial ToxoDB release was not supported by a relational database and thus the site had restricted functionality and little capability to integrate diverse

data types such as gene expression data and single nucleotide polymorphism data (SNPs) with genomic sequence. Since initial publication, ToxoDB has been completely rebuilt using a common architecture similar to another apicomplexan database project, PlasmoDB (5). Both sites, along with CryptoDB, are component sites of ApiDB, the Apicomplexan Bioinformatics Resource Center (6). Many of the new methods of data loading, querying and presentation that are mentioned here have been applied to all of the ApiDB sites to provide a common research platform and facilitate data access among this group of related organisms. ApiDB (<http://apidb.org/>) serves as an ‘umbrella’ site for cross-species comparisons. Researchers can mine for *Toxoplasma* genes at ApiDB directly or via their orthologous relationship(s) to genes in other apicomplexan species.

CONTENT OF THE CURRENT RELEASE

Data

ToxoDB provides access to the genome sequence and annotation of *T. gondii* (ME49 strain) and the genomic sequence of the GT1, VEG and RH (Chr Ia and Chr Ib) strains. Annotation is also available for the apicoplast genome. The current database version (Release 4.2) also contains manual annotation (solicited in the initial genome annotation and entered by users as user comments), ESTs, TIGR Gene Indices clustered ESTs, SAGE tags, SNPs, cosmid and BAC ends, microarray and proteomics studies, all of which have been mapped to the genome (7,8). The database contains the results of automated analyses including gene predictions (using various algorithms), open reading frames (ORFs) greater than 50 aa and protein feature predictions

*To whom correspondence should be addressed: Tel: +1 706 542 6562; Fax: +1 706 542 3582; Email: jkissing@uga.edu

Present address:

Aaron J. Mackey, GlaxoSmithKline, Collegeville, PA, USA

Table 1. Data and analyses that have been integrated into ToxoDB and the number of genes that are impacted

Data type	Data source	Number of genes
Genes	TIGR	8032
Community annotation	Various contributors	1610
Orthologs	Generated from OrthoMCL	4616
GO terms	TIGR; InterPro	3136
EC numbers	TIGR	800
SNPs	John Boothroyd Laboratory; David Roos Laboratory	7322
Microarray	David Roos Laboratory	7664
ESTs	dbEST, TIGR Gene Indices	6080
SAGE tags	TgSAGEDB (14)	6284
Proteomics	Johnathan Wastling Laboratory; John Murray Laboratory	2435
Epitopes	IEDB	10
Metabolic pathways	KEGG Pathway	614

[signal peptides, transmembrane domains, hydrophobicity plots, AA content and InterPro domains (9)], Gene Ontology function predictions, and BLAST similarities to the NCBI non-redundant protein database (Table 1).

In addition, we have used the OrthoMCL algorithm to group genes from *T. gondii* with orthologous genes from 86 other eukaryotic and prokaryotic genomes (10). A mapping of immune epitopes identified in *Toxoplasma* provided by the Immune Epitope Database and Analysis Resource (IEDB) (11) has been integrated. Affymetrix probes mapped to the genome are visible in GBrowse, as are SNPs generated from nucmer alignments of sequences from the GT1, VEG and RH (Chr Ia and Ib) strains against the reference ME49 sequence. Two expression experiments utilizing a *Toxoplasma* Affymetrix array have also been deposited in ToxoDB. Users gain access to these new data types in record pages and by queries using the powerful query interface (see Data-Mining section).

Database architecture

As a part of the complete restructure of the ToxoDB resource, the practice of using flat files as a means of data storage was abandoned in early 2006. We now use GUS 3.5, and load data into an underlying Oracle database in a systematic fashion. GUS (Genomics Unified Schema) is an open source project (www.gusdb.org) with a rich relational schema including sequence annotation, expression data and proteomics using controlled vocabularies and ontologies (12).

ToxoDB also employs the GUS WDK (Web Development Kit, www.gusdb.org/wdk), to access the database from the internet dramatically improving the way the website operates. This transformation has added considerable increased functionality for database users and conforms to the model used by all ApiDB projects, making it possible for us to generate future database releases in short cycles.

DATA-MINING TOOLS

ToxoDB currently provides 40 different queries of the data and several ancillary tools for analyzing, retrieving or viewing the data such as BLAST, Pathway Tools and an installation of the GMOD project Genome Browser (13). The ToxoDB ‘Query & Tools’ page has been restructured to make all queries available at a glance. Most of the individual queries have been reorganized into categories such as ‘Position’, ‘Expression’ and ‘Function’ to make them more intuitive to the average researcher. Enhanced functionality for the queries has also been added. For example, the ToxoDB keyword search has been significantly improved, offering the user control over which fields in the database are searched, including the official annotation, synonyms, user-supplied comments, domain names, BLAST similarities, etc. Many queries, such as ‘Find SNPs based on Gene ID’, now allow a gene ID list as input [either typed (or copied) by hand or uploaded from a file] facilitating analyses on large groups of genes. The results from all queries can be sorted based on various criteria (columns in the returned data set) and users can also add additional criteria for display (e.g. add columns to display protein features, GO annotation, expression characteristics for gene results, etc.) and sort on them as well. Once the appropriate selection of data types to display has been achieved, users can integrate these search results with other search results using the ‘Query History’ page, or the data can be downloaded in multiple formats for further analysis by the researcher (Figure 1).

ToxoDB uses the GBrowse genome browser (www.gmod.org) (13) to display gene models, EST alignments, SNPs, SAGE tags, etc. GBrowse enables visualization of the parasite genome and gene models, custom restriction-site identification, open reading frame identification, and facilitates download of data in various formats. Different data sets or analyses are displayed as individual tracks within the genome browser. There are approximately 50 GBrowse tracks available in the current version of ToxoDB. All genome sequences [ME49, GT1, VEG and RH (Chr Ia, Ib)] are also available in BLAST-searchable databases and for download in FASTA, GenBank and EMBL formats.

ToxoDB users may now register and log in to the site. Doing so enables a researcher to add comments to genes and genomic sequences. It also lets users save query results permanently. Queries in the Query History page can be organized (re-named or deleted) as well as combined with other results (Figure 1). This is a very powerful feature that allows users to refine their results so that precise sets of genes can be discovered.

The results may be downloaded using ToxoDB’s improved reporting facility. It supports summary reports (Excel compatible tab delimited text), GFF, FASTA and a detailed report that includes almost all available data for each gene in the users result table. Use of this facility as well as many others on the site are now described in short video tutorials that are accessible from the database home page.

ToxoDB Release 4.2

ToxoDB Queries and Tools

Identify Genes by:

- Query Availability: ApoDB, CryptDB, PlasmDB, ToxDB
- Genomic Position**: Chromosomal Location, Proximity to Centromeres, Proximity to Telomeres, Non-nuclear Genomes
- Gene Attributes**: Type (e.g. RNA, tRNA), ExonIntron Structure, Keyword, List of IDs, Species, Available Reagents
- Other Attributes**: Similarity/Pattern

Transcript Expression: EST Evidence, mRNA Evidence, Microarray Evidence

Predicted Proteins: Molecular Weight, Isoelectric Point, EC Number, Eptope

Protein Expression: Mass Spec Evidence

Putative Function: GO Term, Metabolic Pathway, 2D Interaction, Predicted Interaction

Similarity/Pattern: Protein Motif, InterPro/Plain Domain, BLAST similarity

Cellular Location: Signal Peptide, Transmembrane Domains, Organellar Compartment, Exported to Host

Evolution: Ortholog/Paralog, Orthology Profile, Homology Profile

Population Biology: SNP, Microsatellites

Identify Genomic Sequences by:: Sequence ID, Species, DNA/Met

Identify SNPs by:: SNP ID, Gene ID, Chromosomal Location

Identify ESTs by:: EST ID, Extent of Gene Overlap, Library

Identify ORFs by:: ORF ID, Mass Spec Evidence, Multi, Chromosomal Location

Tools: Genome browser

View and zoom into regions of genome sequence ranging from chromosomes to nucleotides. Gene annotations, predicted gene models, genetic markers, EST assemblies and other features mapped to the genome are displayed and

Identify Genes based on EST Evidence

EST Library: Tg CAST Tachyzoite cDNA Library, Tg CAST Tachyzoite cDNA Library 2, Tg C001 Tachyzoite, TgME49 1 day, TgME49 7 day, TgME49 77 day, TgME49 Tach, TgME49 inv, TgME49 Tachyzoite, TgME49cDNA, TgRH Tachyzoite, TgRH Tachyzoite Norm 5 cDNA Library, TgRH Tachyzoite Norm 7 cDNA Library, TgRH Tachyzoite Norm 77 cDNA Library, TgRH Tachyzoite cDNA, TgRH Tachyzoite cDNA Library

Identify Genes based on Mass Spec. Evidence

Organism: Toxoplasma gondii

Type: 1-D SDS PAGE¹, 1-D SDS PAGE Soluble Fraction¹, 1-D SDS PAGE Insoluble Fraction¹, MudPIT Insoluble Fraction¹, MudPIT Soluble Fraction¹, Rhopty Fraction¹, Conoid-enriched Fraction², Conoid-depleted Fraction², select all, clear all

Expression of: [1] (Type 1) help, Percentile >= [5] help

Minimum Number of Peptide Sequences: 2

Minimum Number of Spectra: 2

My Query History

ID	Query	Size
3	Rename EST evidence	3946 view download
2	Rename Expression Percentile	3952 view download
1	Rename Mass Spec. Evidence	2156 view download

Combine results: [1 and 2 and 3]

Gene Results

Query: 1 and 2 and 3, Rename | Details: Show | Results: 1298 (showing 1 to 20) | Download | Combine with other results

First 1 2 3 4 5 Next Last To page: [1-488 per page: 20] | Add Column | set Columns | Genomic Sequence ID | Pro strain expression | P strain expression

Gene	Genomic Location	Product Description	Genomic Strand	Pro strain expression	P strain expression
44-m00039	VII: 2,681,601 - 2,693,792 (+)	major surface protein	+	17104.205778672	16173.405718296
59-m00098	VII: 5,304,347 - 5,309,615 (-)	dense granule 1	-	14843.995913723	12862.709611161
42-m00015	X: 919,756 - 918,176 (-)	dense granule 2	-	14843.995913723	12862.709611161
59-m00028	VII: 4,752,930 - 4,754,672 (+)	surface antigen	+	12577.3386446453	12420.998116706
20-m00046	VII: 1,267,900 - 1,270,402 (+)	dense granule 3	+	12577.3386446453	12420.998116706
80-m00191	X: 2,715,216 - 2,720,320 (-)	H4 protein	-	11659.3437966228	11713.3301690007
68-m00022	X: 6,686,399 - 6,689,211 (-)	dense granule 4	-	11659.3437966228	11713.3301690007
78-m00034	V: 1,764,718 - 1,766,335 (+)	microsome pI	+	11628.937207415	10317.704321697
50-m00010	XII: 5,094,435 - 5,096,129 (+)	microsome pI	+	10412.819725915	10412.819725915
80-m00012	X: 4,494,501 - 4,497,830 (-)	microsome pI	-	10412.819725915	10412.819725915
80-m00013	VII: 3,756,099 - 3,760,328 (-)	surface protein	-	10412.819725915	10412.819725915
89-m00024	XI: 736,476 - 739,437 (-)	surface protein	-	10412.819725915	10412.819725915
89-m00025	XI: 2,288,972 - 2,291,723 (+)	SG1-related sequence 2	-	10412.819725915	10412.819725915
44-m00010	VII: 2,288,972 - 2,291,723 (+)	SG1-related sequence 2	-	10412.819725915	10412.819725915
44-m00011	VII: 1,655,336 - 1,657,322 (-)	microsome protein 11 (MC11)	-	10114.559114106	10328.0229343446
49-m00318	VII: 882,209 - 885,249 (-)	hypothetical protein	-	9795.701057684	9795.701057684
33-m00009	X: 9,543,935 - 9,547,209 (+)	M2C-associated protein (MAP)	-	6671.20450306094	6498.8299220534
89-m00021	XI: 5,238,081 - 5,240,945 (-)	superoxide dismutase	-	6671.20450306094	6498.8299220534
80-m00033	X: 3,472,983 - 3,474,993 (-)	glyceraldehyde-3-phosphate dehydrogenase	-	6891.165643475	6891.165643475
78-m00016	V: 1,775,897 - 1,778,347 (-)	elongation factor 1-alpha, putative	-	6891.165643475	6891.165643475

First 1 2 3 4 5 Next Last To page: [1-488 per page: 20] | CO |

76.m00016 elongation factor 1-alpha (putative) Release 4.2

T. gondii protein coding on V from 1,775,897 to 1,778,347

Annotation, **Protein**, **Expression**, **Sequence**, **Data Sources**

Genomic Context Hide

View this sequence in the genome browser

SNPs Summary Show

Annotation

Toxoplasma Paralogs Hide

Find 76.m00016 in the OrthoMCL database

Gene: 83.m00013, **Product**: elongation factor 1-alpha, putative

GO Terms Hide

Is Not	Ontology	GO ID	GO Term Name	Source	Evidence Code
Biological Process	GO:0006412	translation	TIGR	IEA	
Biological Process	GO:0006414	translational elongation	TIGR	IEA	
Biological Process	GO:0006414	translational elongation	Interpro	IEA	
Cellular Component	GO:0005737	cytoplasm	Interpro	IEA	
Cellular Component	GO:0005853	eukaryotic translation elongation factor 1 complex	TIGR	IEA	
Molecular Function	GO:003746	translation elongation factor activity	Interpro	IEA	
Molecular Function	GO:003746	translation elongation factor activity	TIGR	IEA	
Molecular Function	GO:003924	GTPase activity	Interpro	IEA	
Molecular Function	GO:005525	GTP binding	Interpro	IEA	
Molecular Function	GO:005525	GTP binding	TIGR	IEA	

BLASTP Name: 6786420
Description: E = 1.3e-239, 100% identity, gi|95927114|emb|gi|95927114| Toxoplasma gondii elongation factor 1-alpha (EF-1-ALPHA) (TUS029) (Toxoplasma gondii hominis TU502) Expectation: 7.7e-201 % Positive: 86.5 % Positive: 91.0 Coordinates: 1..437

Transmembrane Domains
Wise-2SST little helicity plot

Low Complexity Regions
Low complexity plot

BLASTP Hits

E = 1.3e-239, 100% identity, gi|95927114|emb|gi|95927114| Toxoplasma gondii elongation factor 1-alpha (EF-1-ALPHA) (TUS029) (Toxoplasma gondii hominis TU502)
E = 1.4e-204, 86% identity, gi|95927114|emb|gi|95927114| Toxoplasma gondii elongation factor 1-alpha (EF-1-ALPHA) (TUS029) (Toxoplasma gondii hominis TU502)
E = 7.7e-201, 86% identity, gi|95927114|emb|gi|95927114| Toxoplasma gondii elongation factor 1-alpha (EF-1-ALPHA) (TUS029) (Toxoplasma gondii hominis TU502)
E = 1.6e-204, 85% identity, gi|913732117|emb|gi|913732117| Toxoplasma gondii elongation factor 1-alpha (EF-1-ALPHA) (TUS029) (Toxoplasma gondii hominis TU502)
E = 2.4e-200, 82% identity, TUS029

Protein, **back to top**, **Data Sources**

Protein Features Hide

Mass Spec Peptides (Wassling, et al.)

Mass Spec Peptides (Murray, et al.)

InterPro Domains: TIGR0483

EF_1_alpha, EF_1_alpha1: translation elongation factor EF-1, subunit alpha, PF03143, PF03144

0044777_P-loop containing nucleoside triphosphate hydrolase, PF00035

GTP_EFTU_Elongation Factor Tu GTP binding domain, PF05045, ELONGEFCT_GTP-binding elongation factor signature

Expression, **back to top**

Expression profiling of three archetypal *T. gondii* lineages

Effects of Glucose Starvation

Gene expression analysis

Gene expression analysis

Gene expression analysis

Gene: 76.m00016, **Product**: elongation factor 1-alpha, putative

Gene: 83.m00013, **Product**: elongation factor 1-alpha, putative

Gene: 89-m00016, **Product**: elongation factor 1-alpha, putative

Gene: 9795.701057684, **Product**: elongation factor 1-alpha, putative

Gene: 6671.20450306094, **Product**: elongation factor 1-alpha, putative

Gene: 10412.819725915, **Product**: elongation factor 1-alpha, putative

Gene: 10114.559114106, **Product**: elongation factor 1-alpha, putative

Gene: 6498.8299220534, **Product**: elongation factor 1-alpha, putative

Gene: 6891.165643475, **Product**: elongation factor 1-alpha, putative

Gene: 12577.3386446453, **Product**: elongation factor 1-alpha, putative

Gene: 14843.995913723, **Product**: elongation factor 1-alpha, putative

Gene: 16173.405718296, **Product**: elongation factor 1-alpha, putative

Gene: 12862.709611161, **Product**: elongation factor 1-alpha, putative

Gene: 10317.704321697, **Product**: elongation factor 1-alpha, putative

Gene: 10412.819725915, **Product**: elongation factor 1-alpha, putative

Gene: 10114.559114106, **Product**: elongation factor 1-alpha, putative

Gene: 6498.8299220534, **Product**: elongation factor 1-alpha, putative

Gene: 6891.165643475, **Product**: elongation factor 1-alpha, putative

Gene: 12577.3386446453, **Product**: elongation factor 1-alpha, putative

Gene: 14843.995913723, **Product**: elongation factor 1-alpha, putative

Gene: 16173.405718296, **Product**: elongation factor 1-alpha, putative

Gene: 12862.709611161, **Product**: elongation factor 1-alpha, putative

Gene: 10317.704321697, **Product**: elongation factor 1-alpha, putative

Gene: 10412.819725915, **Product**: elongation factor 1-alpha, putative

Gene: 10114.559114106, **Product**: elongation factor 1-alpha, putative

Gene: 6498.8299220534, **Product**: elongation factor 1-alpha, putative

Gene: 6891.165643475, **Product**: elongation factor 1-alpha, putative

Gene: 12577.3386446453, **Product**: elongation factor 1-alpha, putative

Gene: 14843.995913723, **Product**: elongation factor 1-alpha, putative

Gene: 16173.405718296, **Product**: elongation factor 1-alpha, putative

Gene: 12862.709611161, **Product**: elongation factor 1-alpha, putative

Gene: 10317.704321697, **Product**: elongation factor 1-alpha, putative

Gene: 10412.819725915, **Product**: elongation factor 1-alpha, putative

Gene: 10114.559114106, **Product**: elongation factor 1-alpha, putative

Gene: 6498.8299220534, **Product**: elongation factor 1-alpha, putative

Gene: 6891.165643475, **Product**: elongation factor 1-alpha, putative

Gene: 12577.3386446453, **Product**: elongation factor 1-alpha, putative

Gene: 14843.995913723, **Product**: elongation factor 1-alpha, putative

Gene: 16173.405718296, **Product**: elongation factor 1-alpha, putative

Gene: 12862.709611161, **Product**: elongation factor 1-alpha, putative

Gene: 10317.704321697, **Product**: elongation factor 1-alpha, putative

Gene: 10412.819725915, **Product**: elongation factor 1-alpha, putative

Gene: 10114.559114106, **Product**: elongation factor 1-alpha, putative

Gene: 6498.8299220534, **Product**: elongation factor 1-alpha, putative

Gene: 6891.165643475, **Product**: elongation factor 1-alpha, putative

Gene: 12577.3386446453, **Product**: elongation factor 1-alpha, putative

Gene: 14843.995913723, **Product**: elongation factor 1-alpha, putative

Gene: 16173.405718296, **Product**: elongation factor 1-alpha, putative

Gene: 12862.709611161, **Product**: elongation factor 1-alpha, putative

Gene: 10317.704321697, **Product**: elongation factor 1-alpha, putative

Gene: 10412.819725915, **Product**: elongation factor 1-alpha, putative

Gene: 10114.559114106, **Product**: elongation factor 1-alpha, putative

Gene: 6498.8299220534, **Product**: elongation factor 1-alpha, putative

Gene: 6891.165643475, **Product**: elongation factor 1-alpha, putative

Gene: 12577.3386446453, **Product**: elongation factor 1-alpha, putative

Gene: 14843.995913723, **Product**: elongation factor 1-alpha, putative

Gene: 16173.405718296, **Product**: elongation factor 1-alpha, putative

Gene: 12862.709611161, **Product**: elongation factor 1-alpha, putative

Gene: 10317.704321697, **Product**: elongation factor 1-alpha, putative

Gene: 10412.819725915, **Product**: elongation factor 1-alpha, putative

Gene: 10114.559114106, **Product**: elongation factor 1-alpha, putative

Gene: 6498.8299220534, **Product**: elongation factor 1-alpha, putative

Gene: 6891.165643475, **Product**: elongation factor 1-alpha, putative

Gene: 12577.3386446453, **Product**: elongation factor 1-alpha, putative

Gene: 14843.995913723, **Product**: elongation factor 1-alpha, putative

Gene: 16173.405718296, **Product**: elongation factor 1-alpha, putative

Gene: 12862.709611161, **Product**: elongation factor 1-alpha, putative

Gene: 10317.704321697, **Product**: elongation factor 1-alpha, putative

Gene: 10412.819725915, **Product**: elongation factor 1-alpha, putative

Gene: 10114.559114106, **Product**: elongation factor 1-alpha, putative

Gene: 6498.8299220534, **Product**: elongation factor 1-alpha, putative

Gene: 6891.165643475, **Product**: elongation factor 1-alpha, putative

Gene: 12577.3386446453, **Product**: elongation factor 1-alpha, putative

Gene: 14843.995913723, **Product**: elongation factor 1-alpha, putative

Gene: 16173.405718296, **Product**: elongation factor 1-alpha, putative

Gene: 12862.709611161, **Product**: elongation factor 1-alpha, putative

Gene: 10317.704321697, **Product**: elongation factor 1-alpha, putative

Gene: 10412.819725915, **Product**: elongation factor 1-alpha, putative

Gene: 10114.559114106, **Product**: elongation factor 1-alpha, putative

Gene: 6498.8299220534, **Product**: elongation factor 1-alpha, putative

Gene: 6891.165643475, **Product**: elongation factor 1-alpha, putative

Gene: 12577.3386446453, **Product**: elongation factor 1-alpha, putative

Gene: 14843.995913723, **Product**: elongation factor 1-alpha, putative

Gene: 16173.405718296, **Product**: elongation factor 1-alpha, putative

Gene: 12862.709611161, **Product**: elongation factor 1-alpha, putative

Gene: 10317.704321697, **Product**: elongation factor 1-alpha, putative

Gene: 10412.819725915, **Product**: elongation factor 1-alpha, putative

Gene: 10114.559114106, **Product**: elongation factor 1-alpha, putative

Gene: 6498.8299220534, **Product**: elongation factor 1-alpha, putative

Gene: 6891.165643475, **Product**: elongation factor 1-alpha, putative

Gene: 12577.3386446453, **Product**: elongation factor 1-alpha, putative

Gene: 14843.995913723, **Product**: elongation factor 1-alpha, putative

Gene: 16173.405718296, **Product**: elongation factor 1-alpha, putative

Gene: 12862.709611161, **Product**: elongation factor 1-alpha, putative

Gene: 10317.704321697, **Product**: elongation factor 1-alpha, putative

Gene: 10412.819725915, **Product**: elongation factor 1-alpha, putative

Gene: 10114.559114106, **Product**: elongation factor 1-alpha, putative

Gene: 6498.8299220534, **Product**: elongation factor 1-alpha, putative

Gene: 6891.165643475, **Product**: elongation factor 1-alpha, putative

Gene: 12577.3386446453, **Product**: elongation factor 1-alpha, putative

Gene: 14843.995913723, **Product**: elongation factor 1-alpha, putative

Gene: 16173.405718296, **Product**: elongation factor 1-alpha, putative

Gene: 12862.709611161, **Product**: elongation factor 1-alpha, putative

Gene: 10317.704321697, **Product**: elongation factor 1-alpha, putative

Gene: 10412.819725915, **Product**: elongation factor 1-alpha, putative

Gene: 10114.559114106, **Product**: elongation factor 1-alpha, putative

Gene: 6498.8299220534, **Product**: elongation factor 1-alpha, putative

Gene: 6891.165643475, **Product**: elongation factor 1-alpha, putative

Gene: 12577.3386446453, **Product**: elongation factor 1-alpha, putative

Gene: 14843.995913723, **Product**: elongation factor 1-alpha, putative

Gene: 16173.405718296, **Product**: elongation factor 1-alpha, putative

Gene: 12862.709611161, **Product**: elongation factor 1-alpha, putative

Gene: 10317.704321697, **Product**: elongation factor 1-alpha, putative

Gene: 10412.819725915, **Product**: elongation factor 1-alpha, putative

Gene: 10114.559114106, **Product**: elongation factor 1-alpha, putative

Gene: 6498.8299220534, **Product**: elongation factor 1-alpha, putative

Gene: 6891.165643475, **Product**: elongation factor 1-alpha, putative

Gene: 12577.3386446453, **Product**: elongation factor 1-alpha, putative

Gene: 14843.995913723, **Product**: elongation factor 1-alpha, putative

Gene: 16173.405718296, **Product**: elongation factor 1-alpha, putative

Gene: 12862.709611161, **Product**: elongation factor 1-alpha, putative

Gene: 10317.704321697, **Product**: elongation factor 1-alpha, putative

Gene: 10412.819725915, **Product**: elongation factor 1-alpha, putative

Gene: 10114.559114106, **Product**: elongation factor 1-alpha, putative

Gene: 6498.8299220534, **Product**: elongation factor 1-alpha, putative

Gene: 6891.165643475, **Product**: elongation factor 1-alpha, putative

Gene: 12577.3386446453, **Product**: elongation factor 1-alpha, putative

Gene: 14843.995913723, **Product**: elongation factor 1-alpha, putative

Gene: 16173.405718296, **Product**: elongation factor 1-alpha, putative

Gene: 12862.709611161, **Product**: elongation factor 1-alpha, putative

Gene: 10317.704321697, **Product**: elongation factor 1-alpha, putative

Gene: 10412.819725915, **Product**: elongation factor 1-alpha, putative

Gene: 10114.559114106, **Product**: elongation factor 1-alpha, putative

Gene: 6498.8299220534, **Product**: elongation factor 1-alpha, putative

Gene: 6891.165643475, **Product**: elongation factor 1-alpha, putative

Gene: 12577.3386446453, **Product**: elongation factor 1-alpha, putative

Gene: 14843.995913723, **Product**: elongation factor 1-alpha, putative

Gene: 16173.405718296, **Product**: elongation factor 1-alpha, putative

Gene: 12862.709611161, **Product**: elongation factor 1-alpha, putative

Gene: 10317.704321697, **Product**: elongation factor 1-alpha, putative

Gene: 10412.819725915, **Product**: elongation factor 1-alpha, putative

Gene: 10114.559114106, **Product**: elongation factor 1-alpha, putative

Gene: 6498.8299220534, **Product**: elongation factor

FUTURE DIRECTIONS

The last two years were spent on major infrastructure and design elements for ToxoDB. Our future growth will be in the area of increased data acquisition and integration with existing and future data sets. Specifically, we are planning to load and integrate many expression data sets (RNA expression and protein expression) that are just becoming available. We also expect to load and integrate other array-based data sets such as ChIP on Chip and array CGH. As new data are added, we will be adding additional queries and tools to view these data. An area of significant future development will be improving the ability of users to compare the various different sequenced parasite strains visually and download sequence alignments between them.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the significant contributions of data to the database by members of the *Toxoplasma* research community and the genome sequencing centers. Without their generous contributions, often of pre-publication data, this integrated database resource would not be possible. We also thank the numerous staff and students, both past and present, of the ApiDB-BRC and our research laboratories whose contributions have facilitated the creation and maintenance of this database resource. This project has been funded in whole or in part with Federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under Contract No. HHSN266200400037C. Funding to pay the Open Access publication charges for this article was provided by this contract.

Conflict of interest statement. None declared.

REFERENCES

1. Remington,J.S. and Desmonts,G. (1989) Toxoplasmosis. In Remington,J.S. and Klein,J.O. (eds), *Infectious Diseases of the Fetus and Newborn Infant*. W. B. Saunders, Philadelphia, PA, pp. 89–195.
2. Luft,B.J. and Remington,J.S. (1992) Toxoplasmic encephalitis in AIDS patients. *Clin. Infect. Dis.*, **15**, 211–222.
3. Kissinger,J.C., Gajria,B., Li,L., Paulsen,I.T. and Roos,D.S. (2003) ToxoDB: accessing the *Toxoplasma gondii* genome. *Nucleic Acids Res.*, **31**, 234–236.
4. Khan,A., Taylor,S., Su,C., Mackey,A.J., Boyle,J., Cole,R., Glover,D., Tang,K., Paulsen,I.T. et al. (2005) Composite genome map and recombination parameters derived from three archetypal lineages of *Toxoplasma gondii*. *Nucleic Acids Res.*, **33**, 2980.
5. Bahl,A., Brunk,B., Crabtree,J., Fraunholz,M.J., Gajria,B., Grant,G.R., Ginsburg,H., Gupta,D., Kissinger,J.C. et al. (2003) PlasmoDB. The *Plasmodium* genome resource: a database integrating experimental and computational data. *Nucleic Acids Res.*, **31**, 212–215.
6. Aurrecoechea,C., Heiges,M., Wang,H., Wang,Z., Fischer,S., Rhodes,P., Miller,J., Kraemer,E., Stoeckert,C.J. et al. (2007) ApiDB: integrated resources for the apicomplexan bioinformatics resource center. *Nucleic Acids Res.*, **35**, D427–D430.
7. Hu,K., Johnson,J., Florens,L., Fraunholz,M., Suravajjala,S., DiLullo,C., Yates,J., Roos,D.S. and Murray,J.M. (2006) Cytoskeletal components of an invasion machine – the apical complex of *Toxoplasma gondii*. *PLoS Pathog.*, **2**, e13.
8. Bradley,P.J., Ward,C., Cheng,S.J., Alexander,D.L., Coller,S., Coombs,G.H., Dunn,J.D., Ferguson,D.J., Sanderson,S.J. et al. (2005) Proteomic analysis of rhoptry organelles reveals many novel constituents for host-parasite interactions in *Toxoplasma gondii*. *J. Biol. Chem.*, **280**, 34245–34258.
9. Mulder,N.J., Apweiler,R., Attwood,T.K., Bairoch,A., Bateman,A., Binns,D., Bork,P., Buillard,V., Cerutti,L. et al. (2007) New developments in the InterPro database. *Nucleic Acids Res.*, **35**, D224–D228.
10. Chen,F., Mackey,A.J., Stoeckert,C.J. Jr and Roos,D.S. (2006) OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. *Nucleic Acids Res.*, **34**, D363–D368.
11. Peters,B., Sidney,J., Bourne,P., Bui,H.H., Buus,S., Doh,G., Fleri,W., Kronenberg,M., Kubo,R. et al. (2005) The immune epitope database and analysis resource: from vision to blueprint. *PLoS Biol.*, **3**, e91.
12. Davidson,S., Crabtree,J., Brunk,B.P., Schug,J., Tannen,V., Overton,G.C. and Stoeckert,C.J. Jr (2001) K2/Klesli and GUS: experiments in integrated access to genomic data sources. *IBM Syst. J.*, **40**, 512–531.
13. Stein,L.D., Mungall,C., Shu,S., Caudy,M., Mangone,M., Day,A., Nickerson,E., Stajich,J.E., Harris,T.W. et al. (2002) The generic genome browser: a building block for a model organism system database. *Genome Res.*, **12**, 1599–1610.
14. Radke,J.R., Behnke,M.S., Mackey,A.J., Radke,J.B., Roos,D.S. and White,M.W. (2005) The transcriptome of *Toxoplasma gondii*. *BMC Biol.*, **3**, 26.